IEEE Int. Conf. Neural Networks & Signal Processing
Nanjing, China, December 14-17, 2003

METHODS FOR THE BLIND SIGNAL SEPARATION PROBLEM

Yan Li[1], Peng Wen[2] and David Powers[3]

[1]Department of Mathematics and Computing, Faculty of Sciences; [2]Faculty of Engineering and Surveying, The University of Southern Queensland, QLD 4350, Australia

[3]School of Informatics and Engineering, The Flinders University of Southern Queensland, QLD 4350, Australia
David.Powers@flinders.edu.au

ABSTRACT

This paper classifies and reviews the available algorithms to blind signal separation (BSS) problem. Based on the separation criteria, we broadly divide all the reviewed algorithms into four categories, namely: classical adaptive, higher-order statistics based, information theory based algorithms and others. For algorithms which might fall into more than one category, categorizing is made according to their main features. Most of the algorithms reviewed in this paper are benchmarks in BSS area. Many BSS algorithms use neural networks to perform the learning rules, probably because neural networks are powerful in nonlinear mapping and learning ability.

1. INTRODUCTION

BSS is to recover independent sources given sensor outputs in which the sources have been mixed by unknown channels. Many different approaches to BSS have been attempted by numerous researchers using neural networks, machine learning, higher-order statistics, minimum mutual information, entropy maximization, beam-forming and adaptive noise cancellation, each claiming various degrees of success. This paper reviews the available algorithms to the BSS problem. Based on the separation criteria, we broadly divide all the reviewed algorithms into four categories, namely: classical adaptive, higher order statistics (HOS) based, information theory based algorithms and others. Some algorithms may fall into more than one category, such as Comon's independent component analysis (ICA) [1] which defined a higher-order cumulants contrast function derived from Kullback divergence distance (information theory). In these cases, categorizing is made according to their main features. Most of the algorithms reviewed in this paper are benchmarks in BSS area.

Classical adaptive filter based algorithms are those traditional methods developed in signal processing area. The selected representatives are introduced in section 2. HOS based and information theory based algorithms belong to ICA techniques. Section 3 and 4 will introduce these methods, respectively. Section 5 reviews some other algorithms, which also play very important roles in solving the BSS problem.

2. CLASSICAL ADAPTIVE ALGORITHMS TO BSS PROBLEM

Many early algorithms to the BSS problem were developed in signal processing field.

2.1. Herault and Jutten Algorithm

This is the earliest solution to the blind signal separation problem given by Herault and Jutten [2]. The original algorithm was a two-input/two-output (two sensors/two sources) network called HJ network for BSS.

Herault and Jutten assumed that the source signals are linearly mixed. They proposed a feedback network structure matrix C, which can invert the mixing combinations, $x(t)$. The cascading of the mixing matrix A and the feedback network C, results in the following relation between the estimates of the sources, $y(t)$ and the observations, $x(t)$:

$$y(t) = (I + C)^{-1} A\, x(t) \tag{1}$$

with $diag(C) = \{0, 0, \ldots, 0\}$. An adaptation rule for c_{ij} is given by:

$$\frac{dc_{ij}}{dt} = \alpha\, f(y_1(t))g(y_2(t)),\ \forall i, j\ \ with\ \ i \neq j \tag{2}$$

where $f(.)$ and $g(.)$ are different, odd, non-linear functions and α is a small adaptation or learning constant. The usual choice is $f(x)=x$ and $g(x)=x^3$. Based on this work, various extended versions have been proposed.

2.2. Signal Separation by Output Decorrelation

The objective of output decorrelation is to produce uncorrelated outputs y_i by minimizing a cost function of an arbitrary number of cross-correlation lags (the differences of time instances). The approach uses only second order moments, thus in theory it has limited separating capabilities when compared with HOS methods. However, its simplicity, fast convergence rate and good performance make it particularly suitable for tackling the separation problems that require real-time responses.

[3] suggested an iterative scheme for solving 2x2 linear time invariant problem together with a frequency domain equivalent. For the two-input/two-output case, a cost function, using decorrelation as the criterion, is defined which equals the sum of the squares of the cross-correlations of the

outputs y_1 and y_2. Denoting: $R_{y_1y_2}(l) = E[(y_1(k)y_2(k+l)]$, where l is the lag of the time delay. The cost function C is therefore:

$$C = \sum_{l=0}^{n} R_{y_1y_2}^2(l) \quad (3)$$

Van Gerven and Van Compernolle minimised C using a gradient descent method. Many other methods have been reported using output decorrelation. Li *et al* in [4] proposed a novel iterative algorithm using a recurrent neural network (RNN) for BSS. The cross-correlations of the outputs of the RNN are used as the separation criterion.

2.3. Adaptive Noise Cancellation (ANC)

Widrow *et al* [5] addressed the ANC problem in 1975, which became one of the fundamental signal processing techniques. The concept of ANC is to estimate a desired signal that is corrupted by additive noise or interference. The system analysed in the classic noise cancellation problems is illustrated in Figure 1.

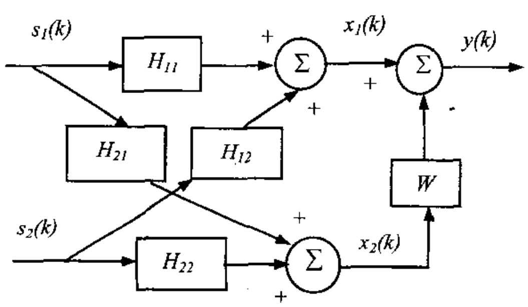

Figure 1 The diagram of classic noise cancellation

The desired source signals $s_1(k)$ and the interfering noise $s_2(k)$ are combined by filters H_{11}, H_{12}, H_{21} and H_{22}, and give the measured signals $x_1(k)$ *and* $x_2(k)$. The idea of ANC is to make $H_{11} \approx 1$ and $H_{21} \approx 0$ as possible at all frequencies. Based on Widrow's work, many modified versions were produced.

3. HIGHER ORDER STATISTICS BASED ALGORITHMS

Most of HOS based algorithms use cumulants, especially the fourth-order cumulants because of their convenient linear properties and their relations to the entropy.

3.1. Burel's Multilayer Neural Network

Based on HJ network, Burel [6] proposed to use a multi-layer neural network to construct the function $g(.)$. The neural network performs signal separation by adapting its weights to minimize a cost function that is designed to make each output have a mean of zero and a variance of unity, and make the outputs independent. The derivation of the cost function is quite mathematically involved. Burel's multi-layer neural network algorithm is also the first algorithm addressing nonlinear mixing model. Good results for separating linear and nonlinear mixed signals were reported in his paper.

3.2. Comon's ICA

Pierre Comon [1] is the first to address the problem of ICA, in a thoroughly mathematical way. The central issue in Comon's ICA is the definition of a contrast function $\Psi(p_s)$ of a n-dimensional probability density function (PDF) p_s for the source vector $\boldsymbol{s}$. Comon showed that the maximisation of the following function leads to source separation:

$$\psi_{ij}(Q) = Cum^2_{iiii}(\tilde{s}) + Cum^2_{jjjj}(\tilde{s}) \quad (4)$$

with respect to an orthogonal matrix transformation Q, for every pair i, j of the n input signals. Cum_{iiii} is the fourth order cumulant of signal s_i. $\tilde{s} = Q\tilde{x}$, where $\tilde{x}$ is the whitened version of x.

3.3. The Alternative Introduced by Cardoso: Serial Updating

Cardoso proposed in [7] a method called serial updating, which is actually closely related to Comon's algorithm. He only considered linear mixing relations, which are inverted by a matrix and updated iteratively. In accordance with the Comon's method, the updates are derived in two steps: whitening of the observed signals and rotation until a criterion function is maximised (or minimised in this case). Contrary to Comon's criterion function of fourth-order cumulants, Cardoso used the fourth-order moments of the output signals y_i to define the criterion function. The merit of Cardoso's method is the linear approximation of the two steps and writing the results as multiplicative errors of previous results, with which the old solutions are updated - Serial Updating.

3.4. Joint Approximate Diagonalization of Eigenmatrices (JADE)

Another important algorithm known as JADE was also proposed by Cardoso [8]. In this algorithm, a cumulant tensor eigenvalue decomposition (EVD) is considered as a preprocessing step. EVD can be viewed as diagonalization. The matrix W diagonalizes $F(M)$ for any M. In other words, $WF(M)W^T$ is diagonal. This is because the matrix F is of a linear combination of the terms of the form $w_iw_i^T$, assuming that the ICA model holds. The diagonality of a matrix $Q = WF(M_i)W^T$ can be measured. A formulation of measuring cumulant matrix with elements $[Q^X(M)]_{ij}$ is defined in [8] as:

$$\phi^{JADE}(y) = \sum_{ijkl \neq iikl} (Q_{ijkl})^2 \quad (5)$$

where ϕ is the contrast function, y is the outputs and i, j, k, l represent the different source signals. The contrast function is effectively a measure of mutual information between the cross-cumulants. Making the cumulants as diagonal as possible can be translated as making the data as independent as possible [8]. The matrix that performs the diagonalization on cumulants can algorithm requires no parameter tuning for good performance. A disadvantage of this approach is that estimating a complete set of fourth-order cumulants requires storage of $O(n^4)$ cumulant matrices.

3.5. Fixed-Point Algorithms

Hyvarinen and his co-workers have introduced a family of fixed-point algorithms. The members of this family are differentiated firstly by the algorithmic approach and secondly by the contrast function used. The key to all the variations is to find independent components by separately maximising the negentropy of each mixture [9]. There are mainly two algorithmic approaches, the symmetric approach and deflation approach, in the fixed-point algorithm class. The symmetric approach uses a modified rule for the update of the unmixing matrix W that enables simultaneous separation of all independent components, whereas the deflation approach updates the columns of W individually, finding the independent components one at a time. Either of these approaches is able to use almost any non-quadratic contrast function to provide estimates of negentropy [9]. The original algorithm use kurtosis, but more recent versions use the hyperbolic tangent, exponential or cubic functions.

The update rule for the deflation method is given by:

$$w^*(k) = C^{-1}E[xg(w(k-1)^T x)] - E[g'(w(k-1)^T x)]w(k-1)$$
$$w(k) = \frac{w^*(k)}{\sqrt{w^*(k)^T C w^*(k)}} \quad (6)$$

where E[.] is the expectation operation, $w^*(k)$ is the complex conjugate of $w(k)$, g can be any suitable non-linear contrast function, with derivative g', and C is the covariance matrix of the mixes, x.

4. INFORMATION THEORY BASED ALGORITHMS

Information theory based algorithms mainly include entropy maximization and mutual information minimization.

4.1. Bell and Sejnowski's Information Maximization

The information maximisation algorithm (often known as Infomax) developed by Bell and Sejnowski [10] catalysed a surge of interest in using information theory to perform BSS. They proposed in 1995 an adaptive learning algorithm that maximizes the information passed through a neural network, and showed that the neural network is capable of resolving the independent components in the inputs, that is, the neural network can perform independent component analysis. They addressed that maximizing the joint entropy $H(y)$ of the output of a neural processor can approximately minimize the mutual information among the output components. A single layer feedforward neural network was used to implement the separation. The learning rule of the Infomax is.

translated to perform separation on the mixed data. JADE

$$\Delta W \propto [W^T]^{-1} + (1-2y)x^T \quad (7)$$
$$\Delta W_0 \propto 1 - 2y$$

The sigmoid function is used as the transfer function in the paper. The algorithm successfully separated up to ten sources. However, the Infomax algorithm is limited to separate sources with super-Gaussian distributions.

4.2. The Extended Infomax Algorithm

The purpose of the extended Infomax algorithm [11] is to provide a simple learning rule with a fixed nonlinearity that can separate sources with a variety of distributions. [11] generalised the learning rule to sources with either sub- or super- Gaussian distributions. The learning algorithm is

$$\Delta W = [I - K \cdot \tanh(y) \cdot y^T - yy^T] \cdot W$$
$$K_i = 1, \quad super-Gaussian \quad (8)$$
$$K_i = -1, \quad sub-Gaussian$$

here: $\tanh(u) = \frac{e^u + e^{-u}}{e^u - e^{-u}}$ is the hyperbolic tangent; K_i are elements of the n-dimensional diagonal matrix K.

4.3. Shun-Ichi Amari's Natural Gradient

Amari *et al* [12] used a different approach which resulted in the same learning rule as Bell and Sejnowski's, and then it was further extended by performing descent on the natural gradient. The derivation given by Amari used the Kullback-Leibler distance (measure of PDF similarity) as the starting point. If the joint entropy of the outputs is the same as the product of the individual entropies, then the outputs are independent. One way to do so is to minimize the Kullback-Leibler distance between the PDF of the output vector and the product of the PDFs of the individual outputs with respect to the unmixing matrix. Then the natural gradient learning rule is

$$\Delta W \propto [I - f(y) \cdot y^T] \cdot W \quad (9)$$

This rule speeds up the algorithm by removing the computationally expensive matrix inversion in Bell and Sejnowski's algorithm.

5. OTHER ALGORITHMS

There are many other algorithms having been played very important roles in solving BSS problem.

5.1. The Approaches in Frequency Domain

Unlike the previous introduced algorithms, these approaches operate in the frequency domain instead of the time domain. The inputs are transformed to the frequency domain using Fourier Transform. In frequency domain, Smaragdis [13] used a bank of separation networks to separate the Fourier coefficients, which are linearly mixed. The separated bins are then transformed back to the time domain to give back the original sources. Moving to the frequency domain allows us to implement the separation filters using fast convolution algorithms which are well developed in signal processing area.

A problem for the approaches in frequency domain is to obtain a uniform permutation for the unmixing matrices over all frequency bins. In the time domain, arbitrary permutation and scaling of the outputs are allowed in BSS algorithms. However, in the frequency domain processing, different permutation at different frequencies lead to mixing of signals in the final outputs. Also different scaling at different frequencies leads to distortion of the frequency spectrum of the output signals.

5.2. Nonlinear ICA

Researchers have very recently started addressing the ICA formulation to nonlinear mixing models. There are a large group of nonlinear ICA papers in ICA'2001 conference.

The proposed nonlinear ICA methods can be roughly divided into two classes of approaches. The first class of methods is an obvious extension to the linear ICA models where nonlinear mixing models are added to the linear models and the task is to find the inverse of the linear models as well as the inverse of the nonlinear models [11]. The nonlinearities are often parameterised allowing limited flexibility. The second class of methods to the nonlinear ICA problem using self-organising maps (SOM) were proposed by Lin and Cowan [14] and several other research groups. Their approaches were more flexible to some extent parameter-free which allows greater freedom of nonlinear representation. SOM constitutes a class of vector quantizers that impose a prescribed topological order on the reference vectors.

5.3. Semi-Blind Signal Separation

In certain applications some source signals involved in the signal mixing process are known, such as in a teleconferencing setup, where the loudspeakers' signals are accessible. This situation is referred to as spatially semi-blind source separation. For some reason, semi-BSS has remained a rather limited and narrow research effort. Only very a few papers claim to address such general problems. It is suspected that the known information, such as the characteristics of microphones and locations of the sources and microphones, is difficult to be determined in real environment and often subject to some conditions.

6. CONCLUSION

This paper classifies and reviews the available algorithms to the BSS problem. Many algorithms discussed in the paper are benchmarks in BSS area.

7. REFERENCES

[1]. P. Comon, "Independent component analysis - a new concept?", *Signal Processing*, vol. 36, pp. 287-314, 1994.

[2]. J. Herault and C. Jutten, "Space or time adaptive signal processing by neural network models", In J. S. Denker (ed), editor, *Neural Networks For Computing: AIP Conference Proceedings 151*, American Institute for Physics, New York, 1986.

[3]. D. Van Compernolle and S. Van Gerven, "On use of decorrelation in scalar signal separation", In *IEEE, ICASSP*, Vol. 3, pp.57-60, 1994.

[4]. Y. Li, D. Powers and K. Pope, "Blind signal separation using recurrent neural networks", *International Journal of Knowledge-based Intelligent Engineering Systems*, Vol. 7, No. 2, pp. 62-69, 2003.

[5]. B. Widrow, J. R. Glover, J. M., McCool, *et al*, "Adaptive noise canceling: Principles and applications", *IEEE Proc.*, Vol. 63, No. 12, pp. 1692-1716, 1975.

[6]. G. Burel, "Blind separation of sources – a nonlinear neural algorithm", *Neural Networks*, Vol. 5, No. 6, pp. 937-947, 1992.

[7]. J. F. Cardoso, "Iterative techniques for blind separation using only fourth-order cumulants, signal processing IV-theories and applications", *Proceedings of EUSIPCO-92*, Sixth European Signal Processing Conference, Vol. 2, pp. 739-742, 1992.

[8]. J. F. Cardoso and A. Souloumiac, "Blind beamforming for non-Gaussian signals", *IEE Proceeding Part F*, Vol. 140, No. 6, pp. 362-370, 1993.

[9]. A. Hyvarinen, "A family of fixed-point algorithms for independent component analysis", In *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'97)*, pp. 3917-1920, Munich, Germany, 1997.

[10].. A. J. Bell and T. J. Sejnowski, "An information-maximisation approach to blind separation and blind deconvolution", *Neural Computation*, Vol. 7, No. 6, 1004-1034, 1995.

[11]. T. W. Lee, *Independent component analysis-theory and application*, Dordrecht, The Netherlands: Kluwer, 1998.

[12]. S. A. Amari, A. Cichocki, and H. H. Yang, "A new learning algorithm for blind signal separation", *Advances in Neural Information Processing Systems*, Vol. 8, pp. 752-763, MIT Press, Cambridge, MA, 1996.

[13]. P. Smaragdis, *Information theoretic Approaches to source separation*, Master's Thesis, MIT, Cambridge, MA, 1997.

[14]. J. Lin, D. Grier, and J. Cowan, "Faithful representation of separable distributions", *Neural Computation*, Vol. 9, pp.1305-1320, 1997.